SIAM J. CONTROL OPTIM.
Vol. 36, No. 2, pp. 742–759, March 1998

013

A NEW VALUE ITERATION METHOD FOR THE AVERAGE COST DYNAMIC PROGRAMMING PROBLEM*

DIMITRI P. BERTSEKAS†

Abstract. We propose a new value iteration method for the classical average cost Markovian decision problem, under the assumption that all stationary policies are unichain and that, furthermore, there exists a state that is recurrent under all stationary policies. This method is motivated by a relation between the average cost problem and an associated stochastic shortest path problem. Contrary to the standard relative value iteration, our method involves a weighted sup-norm contraction, and for this reason it admits a Gauss–Seidel implementation. Computational tests indicate that the Gauss–Seidel version of the new method substantially outperforms the standard method for difficult problems.

Key words. dynamic programming, average cost, value iteration

AMS subject classifications. 90C35, 49L20

PII. S0363012995291609

1. Introduction.

We consider a controlled discrete-time dynamic system with n states, denoted $1,\dots,n$. At each time, if the state is i, a control u is chosen from a given finite constraint set $U(i)$, and the next state is j with given probability $p_{ij}(u)$. An admissible policy is a sequence of functions from states to controls, $\pi = \{\mu_0, \mu_1, \dots\}$, where $\mu_k(i) \in U(i)$ for all i and k. The average cost corresponding to π and initial state i is

$$J_\pi(i) = \limsup_{N\to\infty} \frac{1}{N} E\left\{ \sum_{k=0}^{N-1} g\big(x_k, \mu_k(x_k)\big) \,\middle|\, x_0 = i \right\},$$

where x_k is the state at time k and g is a given cost function. A *stationary policy* is an admissible policy of the form $\pi = \{\mu, \mu, \dots\}$, and its corresponding cost function is denoted by $J_\mu(i)$. For brevity, we refer to $\{\mu, \mu, \dots\}$ as the stationary policy μ. We want to solve the classical problem of finding an optimal policy, that is, an admissible policy π such that $J_{\pi^*}(i) = \min_\pi J_\pi(i)$ for all i.

A stationary policy is called *unichain* if it gives rise to a Markov chain with a single recurrent class. Throughout the paper, we assume the following.

Assumption 1: All stationary policies are unichain. Furthermore, state n is recurrent in the Markov chain corresponding to each stationary policy.

It is well known that under Assumption 1, the optimal cost $J^*(i)$ has a common value for all initial states, which is denoted by λ^*,

$$J^*(i) = \lambda^*, \qquad i = 1, \dots, n.$$

Furthermore, λ^* together with a differential cost vector $h = \big(h(1), \dots, h(n)\big)$ satisfies Bellman's equation

$$\lambda^* + h(i) = \min_{u\in U(i)} \left[g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h(j) \right], \qquad i = 1, \dots, n. \tag{1}$$

*Received by the editors September 11, 1995; accepted for publication (in revised form) February 4, 1997. This research was supported by NSF grant 9300494-DMI.
http://www.siam.org/journals/sicon/36-2/29160.html

†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 (bertseka@lids.mit.edu).

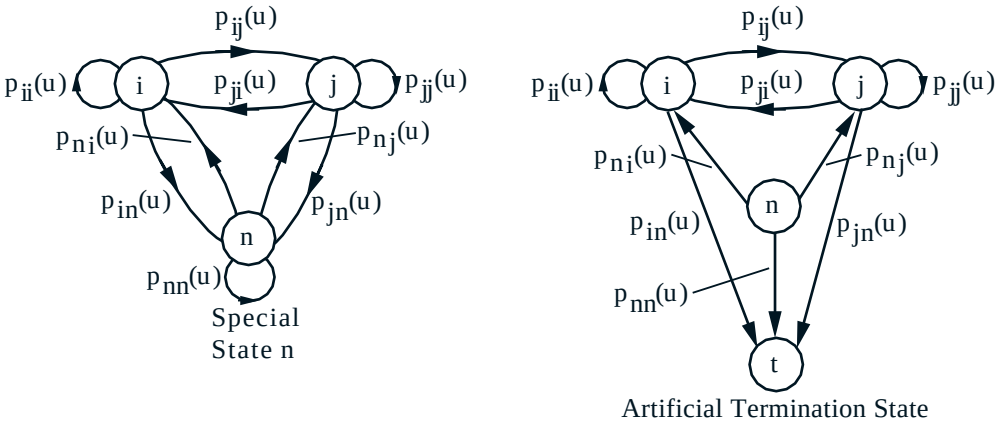

FIG. 1. *Transition probabilities for an average cost problem and its associated stochastic shortest path problem. The latter problem is obtained by introducing, in addition to* $1, \ldots, n$, *an artificial termination state* t *to which we move from each state* i *with probability* $p_{in}(u)$, *by setting all transition probabilities* $p_{in}(u)$ *to* 0, *and by leaving unchanged all other transition probabilities.*

In addition, a stationary policy μ is optimal if and only if $\mu(i)$ attains the minimum in the above equation for all i. These results can be shown under the assumption that all stationary policies are unichain, without requiring the additional condition that there is a common recurrent state to all stationary policies. However, for the methods of this paper, the existence of a common recurrent state is essential, at least for the purposes of analysis. From the computational point of view, the existence of a common recurrent state is less significant, as long as all stationary policies are unichain. One may modify the problem so that Assumption 1 holds by adding a very small positive ϵ to all transition probabilities of the form $p_{in}(u)$. The effect on the average cost per stage of each stationary policy will be $O(\epsilon)$.

Under Assumption 1 we can make an important connection of the average cost problem with an associated stochastic shortest path problem, which has been the basis for a recent textbook treatment of the average cost problem [Ber95, Vol. I, section 7.4]. This problem is obtained by leaving unchanged all transition probabilities $p_{ij}(u)$ for $j \neq n$, by setting all transition probabilities $p_{in}(u)$ to 0, and by introducing an artificial cost-free and absorbing termination state t to which we move from each state i with probability $p_{in}(u)$; see Fig. 1. The expected stage cost at state i of the stochastic shortest path problem is $g(i,u) - \lambda$, where λ is a scalar parameter. Let $h_{\mu,\lambda}(i)$ be the cost of stationary policy μ for this stochastic shortest path problem, starting from state i; that is, $h_{\mu,\lambda}(i)$ is the total expected cost incurred starting from state i up to reaching the termination state t. We refer to this problem as λ-SSP. Let $h_\lambda(i) = \min_\mu h_{\mu,\lambda}(i)$ be the corresponding optimal cost of the λ-SSP. Then the following can be shown (see Fig. 2).

(a) For all μ and λ, we have

$$h_{\mu,\lambda}(i) = h_{\mu,\lambda_\mu}(i) + (\lambda_\mu - \lambda)N_\mu(i), \qquad i = 1, \ldots, n, \tag{2}$$

where $N_\mu(i)$ is the average number of steps required to reach n under μ starting from state i, and λ_μ is the average cost corresponding to μ.

(b) The functions

$$h_\lambda(i) = \min_\mu h_{\mu,\lambda}(i), \qquad i = 1, \ldots, n, \tag{3}$$

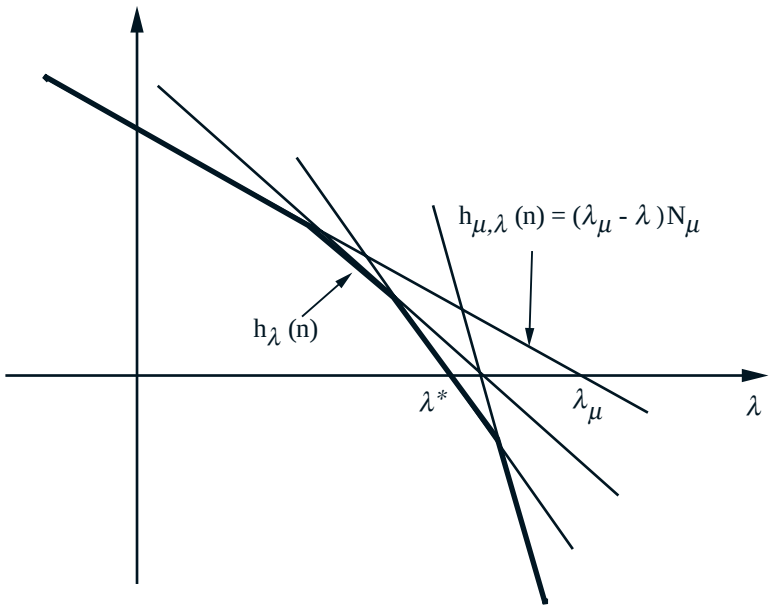

FIG. 2. *Relation of the costs of stationary policies in the average cost problem and the associated stochastic shortest path problem.*

are concave, monotonically decreasing, and piecewise linear as functions of λ, and

$$h_\lambda(n) = 0 \qquad \text{if and only if} \qquad \lambda = \lambda^*. \tag{4}$$

Furthermore, the vector h_{λ^*} satisfies Bellman's equation (1) together with λ^*.

From Fig. 2, it can be seen that λ^* can be obtained by a one-dimensional search procedure that brackets λ^* within a sequence of nested and diminishing intervals; see [Ber95, Vol. II, Fig. 4.5.2]. This method is probably inefficient because it requires the (exact) solution of several λ-SSPs, corresponding to several different values of λ. An alternative method, which is also inefficient because it requires the exact solution of several λ-SSPs, is to update λ by an iteration of the form

$$\lambda^{k+1} = \lambda^k + \gamma^k h_{\lambda^k}(n), \tag{5}$$

where γ^k is a positive stepsize parameter. This iteration is motivated by Fig. 2, where it is seen that $\lambda < \lambda^*$ (or $\lambda > \lambda^*$) if and only if $h_\lambda(n) > 0$ (or $h_\lambda(n) < 0$, respectively). Indeed, it can be seen from Fig. 2 that the sequence $\{\lambda^k\}$ generated by (5) converges to λ^* provided the stepsize γ^k is the same for all iterations and does not exceed the threshold value $1/\max_\mu N_\mu(n)$. Such a stepsize is sufficiently small to guarantee that the difference $\lambda - \lambda^*$ does not change sign during the algorithm (5). Note that each λ-SSP can be solved by value iteration, which has the form

$$h^{k+1}(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n-1} p_{ij}(u) h^k(j) \right] - \lambda, \qquad i = 1, \ldots, n, \tag{6}$$

with λ kept fixed throughout the value iteration method.

In this paper we propose algorithms based on the λ-SSP, which are more efficient than the algorithms mentioned above. In particular, we change λ during the value iteration process (6) by using an iteration of the form (5), but with $h_{\lambda^k}(n)$ replaced by an approximation, the current value iterate $h^{k+1}(n)$. Such an algorithm may be viewed as a *value iteration algorithm for a slowly varying stochastic shortest path problem*. It has the form

$$h^{k+1}(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n-1} p_{ij}(u) h^k(j) \right] - \lambda^k, \qquad i = 1, \ldots, n, \tag{7}$$

$$\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n), \tag{8}$$

where γ^k is a positive stepsize. We prove convergence of this method for the case where γ^k is a sufficiently small constant. Convergence can also be similarly proved for a variety of other stepsize rules.

Our method should be contrasted with the standard relative value iteration method for average cost problems due to [Whi63], which takes the form (see, e.g., [Ber95], [Put94])

$$\lambda^{k+1} = \min_{u \in U(n)} \left[g(n,u) + \sum_{j=1}^{n} p_{nj}(u) h^k(j) \right], \tag{9}$$

$$h^{k+1}(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h^k(j) \right] - \lambda^{k+1}, \qquad i = 1, \ldots, n. \tag{10}$$

If we use (7) to write iteration (8) in the equivalent form

$$\lambda^{k+1} = (1 - \gamma^k)\lambda^k + \gamma^k \min_{u \in U(n)} \left[g(n,u) + \sum_{j=1}^{n-1} p_{nj}(u) h^k(j) \right],$$

we see that if $\gamma^k = 1$ for all k, the new value iteration (7)–(8) becomes similar to the standard value iteration (9)–(10): the updating formulas are the same in both methods (because we have $h^k(n) = 0$ for all $k \geq 1$ in the iteration (9)–(10)), but the order of updating λ is just reversed relative to the order of updating h. Despite the similarity of the two methods, the proof of convergence of the standard method (9)–(10) (as given, for example, in [Ber95, Vol. II] or [Put94]) does not seem to be applicable to the new method. The line of proof given in the next section is substantially different, and makes essential use of Assumption 1 and the connection with the stochastic shortest path problem. Furthermore, one can construct examples where Assumption 1 is violated because state n is transient under some stationary policy, and where the new method (7)–(8) does not converge while the known method (9)–(10) converges. Conversely, it can be seen that the standard aperiodicity assumption required for convergence of the standard method (9)–(10) (see, e.g., [Ber95], [Put94]) is not needed for the new method. We note also that there is a variant of the standard method (9)–(10) that does not require an aperiodicity assumption and involves interpolations

between h^k and h^{k+1} according to a stepsize parameter (see [Sch71, [Pla77], [Var78], [PBW79], [Put94], [Ber95]). However, the new method does not seem as closely related to this variant.

A significant improvement in the algorithm, which guarantees that bounded iterates will be generated for any choice of stepsize, is to calculate upper and lower bounds on λ^* from iteration (7) and then modify iteration (8) to project the iterate $\lambda^k + \gamma^k h^k(n)$ on the interval of the bounds. In particular, based on the Odoni bounds [Odo69] for the relative value iteration method (see, e.g., [Ber95, Vol. II, p. 209], it can be seen that

$$\underline{\beta}^k \leq \lambda^* \leq \overline{\beta}^k,$$

where

$$\underline{\beta}^k = \lambda^k + \min\left[\min_{i \neq n}\big[h^{k+1}(i) - h^k(i)\big],\, h^{k+1}(n)\right], \tag{11}$$

$$\overline{\beta}^k = \lambda^k + \max\left[\max_{i \neq n}\big[h^{k+1}(i) - h^k(i)\big],\, h^{k+1}(n)\right]. \tag{12}$$

Thus we may replace the iteration $\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n)$ (cf. (8)) by

$$\lambda^{k+1} = \Pi_k\big[\lambda^k + \gamma^k h^{k+1}(n)\big], \tag{13}$$

where $\Pi_k[c]$ denotes the projection of a scalar c on the interval

$$\left[\max_{m=0,\ldots,k} \underline{\beta}^m,\ \min_{m=0,\ldots,k} \overline{\beta}^m\right]. \tag{14}$$

We note that the issue of stepsize selection is crucial for the success of our algorithm. In particular, if γ^k is a chosen constant but very small, or diminishing at the rate of $1/k$ (as is common in stochastic approximation algorithms), then λ changes slowly relative to h, and the iteration (8) essentially becomes identical to iteration (5) but with a very small stepsize, which leads to slow convergence. On the other hand, if γ^k is too large, λ^k will oscillate and diverge. One may keep the stepsize γ^k constant at a value found by trial and error, but there are some better alternatives. One possibility that has worked quite reliably and efficiently in our tests is to start with a fairly large γ^k and gradually diminish it if $h^k(n)$ changes sign frequently; for example, we may use

$$\gamma^k = m(\hat{k})\gamma, \tag{15}$$

where

(a) γ is the initial stepsize (a positive constant),

(b) $m(\hat{k})$ is a decreasing function of $\hat{k}$, which is defined as the number of indexes $t \leq k$ such that $h^{t-1}(n)h^t(n) < 0$ and $|h^t(n)|$ is greater than some fixed threshold θ.

Examples of functions $m(\cdot)$ that we tried are

$$m(\hat{k}) = \frac{1}{\hat{k}+1} \tag{16a}$$

and

$$m(\hat{k}) = \xi^{\hat{k}}, \tag{16b}$$

where ξ is a fixed scalar from the range $(0,1)$, so that γ^k is decreased by a factor ξ each time $\hat{k}$ is incremented. Our experience indicates that it is best to choose the initial stepsize γ in the range $[1,5]$. Typically, the stepsize is reduced quickly according to (15) to an appropriate level (which depends on the problem) and then stays constant for the remaining iterations. In our experiments, we have used the preceding choices of γ^k with $\gamma = 1$, $\xi = 0.95$, and $\theta = 1$.

The motivation for our method is that value iteration for stochastic shortest path problems involves a contraction. In particular, consider the mapping $F : \Re^n \to \Re^n$ with components given by

$$F_i(h) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n-1} p_{ij}(u) h(j) \right], \qquad i = 1, \ldots, n.$$

It is known (see, e.g., [BeT89, p. 325] or [Tse90]) that, under Assumption 1, F is a contraction mapping with respect to some weighted sup-norm; that is, for some positive scalars $v_1, \ldots, v_n$, and some scalar $\alpha \in (0,1)$, we have

$$\max_{i=1,\ldots,n} \frac{|F_i(h) - F_i(\overline{h})|}{v_i} \leq \alpha \max_{i=1,\ldots,n} \frac{|h(i) - \overline{h}(i)|}{v_i} \qquad \forall\ h,\ \overline{h} \in \Re^n. \tag{17}$$

Note here that while there is coupling between the iteration of h as per (7) and the iteration for λ as per (8), the latter iteration can be made much slower than the former through the use of the stepsize γ, so that the weighted sup-norm contraction character of the iteration (7) is preserved. Furthermore, even when the stepsize γ is not small, the contraction property of F is analytically convenient, as will be seen, for example, in the analysis of section 3. By contrast, the standard relative value iteration method (9)–(10) does not involve a weighted sup-norm contraction, and in fact it may not involve a contraction of any kind, unless an additional aperiodicity assumption on the Markov chains corresponding to the stationary policies is imposed. We speculate that the sup-norm contraction structure may be helpful in other contexts, beyond those discussed in this paper; for example, in Q-learning (stochastic approximation) variants of the method and when parallel asynchronous variations are considered. In fact, an analysis of Q-learning variants of our method that admit a parallel asynchronous implementation is the subject of a forthcoming report [ABB97].

The new method (7)–(8) can be viewed as a Jacobi type of method, since all the components of h are simultaneously updated. A particularly interesting fact is that the weighted sup-norm contraction property of the mapping F can also be exploited to construct valid Gauss–Seidel variants, where the components of h are updated sequentially in some order. In particular, the method of proof of the next section can be used to show convergence for the Gauss–Seidel version of the method, given by

$$h^{k+1}(i) = G_i(h^k, \lambda^k), \qquad i = 1, \ldots, n,$$

$$\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n),$$

where $G : \Re^{n+1} \to \Re^n$ is the Gauss–Seidel mapping based on F, having components given by

$$G_1(h, \lambda) = \min_{u \in U(1)} \left[g(1,u) + \sum_{j=1}^{n-1} p_{1j}(u) h(j) \right] - \lambda,$$

$$G_i(h,\lambda) = \min_{u\in U(i)} \left[g(i,u) + \sum_{j=1}^{i-1} p_{ij}(u) G_j(h,\lambda) + \sum_{j=i}^{n-1} p_{ij}(u) h(j) \right] - \lambda, \qquad i = 2,\ldots,n.$$

By contrast, we do not know of any convergent Gauss–Seidel version of the standard value iteration (9)–(10). In fact, simple counterexamples show that the straightforward Gauss–Seidel variant of the standard method may diverge.

Note that the Odoni bounds (11)–(12) are not available when the Gauss–Seidel variant is used. However, it is still possible to use the projection (13)–(14) by performing once in a while (say, every 10 iterations) the regular (Jacobi) version (7)–(8) of the method, and obtain corresponding Odoni bounds that can be used for projection at all subsequent iterations. This device proved to be very effective in our experiments.

Regarding a theoretical comparison of the performance of the new methods and the standard method, it can be seen with simple examples that neither type of method dominates the other. Suppose, for instance, that there is only one policy and that the corresponding transition probability matrix is

$$\begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon & \epsilon \end{pmatrix},$$

where ϵ is a scalar from $[0,1]$. Then both methods (7)–(8) and (9)–(10) become linear iterations, and their rate of convergence is governed by the eigenvalues of the corresponding iteration matrix. The eigenvalues corresponding to the standard relative value iteration (9)–(10) can be shown to be 0 and $1-2\epsilon$, so that the method converges very fast for $\epsilon \sim 1/2$ and slowly for $\epsilon \sim 0$ or $\epsilon \sim 1$. It can also be verified that, for a constant but well-chosen value of γ, the eigenvalue structure of the new value iteration method (7)–(8) is worse than the one for the standard method for $\epsilon \sim 1/2$, more favorable for $\epsilon \sim 0$, and comparably unfavorable for $\epsilon \sim 1$.

Our limited computational experiments also indicate that the Jacobi version (7)–(8) of the new method, when properly implemented with the adaptive stepsize rule (15) and the projection scheme of (11)–(14), is competitive with the relative value iteration method of (9)–(10). There are problems where one method outperforms the other and vice versa. When the initial stepsize γ in (15) is equal to 1, the performance of the two methods appears to be quite similar for many problems (see, e.g., Tables 2 and 3 in section 4). On the other hand, our computational results indicate that the Gauss–Seidel variant of the new method substantially outperforms the standard method for relatively difficult problems. This is not surprising, since Gauss–Seidel methods are known to have better performance than their Jacobi counterparts when a weighted sup-norm contraction is involved. Both the standard method and the new methods can be very slow on unfavorably structured problems. This is to be expected, since these methods exhibit convergence rate behavior similar to linear iterations and are subject to ill-conditioning.

The paper is organized as follows. In the next section we prove a convergence result for the Jacobi version of the new method. In section 3 we extend this result to apply to the Gauss–Seidel variant. The method of proof can also be used to prove convergence of a variety of other variants involving different orders of updating the components of the vector h, as well as asynchronous versions. All this flexibility is possible thanks to the weighted sup-norm contraction property of the mapping F. Finally, in section 4 we describe some of our computational experience. In particular, we compare the standard method (9)–(10) with implementations of the Jacobi and Gauss–Seidel versions of our method, which involve an adaptive stepsize rule like the

one of (15) and the projection scheme of (11)–(14). We find that the Gauss–Seidel method outperforms the other methods on the more difficult problems.

2. Convergence analysis. We now investigate the convergence of the new value iteration algorithm. For convenience, let us denote by $\|\cdot\|$ the weighted sup-norm with respect to which the contraction property of (17) holds; that is,

$$\|h\| = \max_{i=1,\ldots,n} \frac{|h(i)|}{v_i} \qquad \forall\ h \in \Re^n.$$

Let us also normalize the vector v so that its last coordinate is equal to 1; that is,

$$v_n = 1.$$

Note that since h_λ is the optimal cost vector of the λ-SSP, we have that h_λ is the unique fixed point of the contraction mapping $F(h) - \lambda e$; that is,

$$h_\lambda = F(h_\lambda) - \lambda e \qquad \forall\ \lambda \in \Re. \tag{18}$$

By writing for all stationary policies μ, states i, and scalars λ and λ',

$$h_{\mu,\,\lambda}(i) = h_{\mu,\,\lambda'}(i) + N_\mu(i)(\lambda' - \lambda),$$

and by using the definition $h_\lambda(i) = \min_\mu h_{\mu,\lambda}(i)$, we obtain the following relation:

$$h_{\lambda'}(i)+\underline{N}(\lambda'-\lambda) \le h_\lambda(i) \le h_{\lambda'}(i)+\overline{N}(\lambda'-\lambda) \qquad \forall\, i = 1,\ldots,n, \text{ and } \lambda,\, \lambda' \in \Re, \tag{19}$$

where $\underline{N}$ and $\overline{N}$ are the positive scalars

$$\underline{N} = \min_\mu \min_{i=1,\ldots,n} N_\mu(i), \qquad \overline{N} = \max_\mu \max_{i=1,\ldots,n} N_\mu(i).$$

We can write (19) in the equivalent form

$$N|\lambda' - \lambda| \le |h_\lambda(i) - h_{\lambda'}(i)| \le \overline{N}|\lambda' - \lambda|, \qquad \forall\ i \text{ and } \lambda, \lambda' \in \mathcal{R}. \tag{20}$$

We can interpret $\underline{N}$ and $\overline{N}$ as uniform lower and upper bounds on the slope of the piecewise linear function $h_\lambda(i)$, viewed as a function of λ (see Fig. 2).

The following is our main result.

PROPOSITION 1. *There exists a positive scalar $\overline{\gamma}$ such that if*

$$\underline{\gamma} \le \gamma^k \le \overline{\gamma} \tag{21}$$

for some positive scalar $\underline{\gamma}$ and all k, the sequence (h^k, λ^k) generated by iteration (7), (8) converges to $(h_{\lambda^*}, \lambda^*)$ at the rate of a geometric progression.

Proof. We will show that there exists a threshold value $\overline{\gamma} > 0$ and a continuous function $c(\gamma)$ with $0 \le c(\gamma) < 1$ for all $\gamma \in (0, \overline{\gamma}]$ such that for any $B > 0$, the relations

$$\|h^k - h_{\lambda^k}\| \le B \quad \text{and} \quad |\lambda^k - \lambda^*| \le \frac{B}{\underline{N}} \tag{22}$$

imply that

$$\|h^{k+1} - h_{\lambda^{k+1}}\| \le c(\gamma^k)B \quad \text{and} \quad |\lambda^{k+1} - \lambda^*| \le \frac{c(\gamma^k)B}{\underline{N}}. \tag{23}$$

This implies that for a stepsize sequence satisfying the assumptions of the proposition, the sequence $|\lambda^k - \lambda^*|$ converges to zero at the rate of a geometric progression, and the same is true of the sequence $\|h^k - h_{\lambda^k}\|$. Since, using (20), we have

$$\|h^k - h_{\lambda^*}\| \le \|h^k - h_{\lambda^k}\| + \|h_{\lambda^k} - h_{\lambda^*}\| \le \|h^k - h_{\lambda^k}\| + O\big(|\lambda^k - \lambda^*|\big),$$

we see that $\|h^k - h_{\lambda^*}\|$ also converges to zero at the rate of a geometric progression.

We first show two preliminary relations. We have, using (18),

$$\begin{aligned}\|h_{\lambda^{k+1}} - h_{\lambda^k}\| &= \|F\big(h_{\lambda^{k+1}}\big) - \lambda^{k+1}e - F\big(h_{\lambda^k}\big) + \lambda^k e\| \\ &\le \|F\big(h_{\lambda^{k+1}}\big) - F\big(h_{\lambda^k}\big)\| + \|(\lambda^{k+1} - \lambda^k)e\| \\ &\le \alpha\|h_{\lambda^{k+1}} - h_{\lambda^k}\| + |\lambda^{k+1} - \lambda^k|\|e\|.\end{aligned}$$

Thus

$$\|h_{\lambda^{k+1}} - h_{\lambda^k}\| \le \frac{\|e\|}{1-\alpha}|\lambda^{k+1} - \lambda^k|. \tag{24}$$

Also, by subtracting the relations

$$h^{k+1}(n) = F_n(h^k) - \lambda^k,$$

$$h_{\lambda^k}(n) = F_n(h_{\lambda^k}) - \lambda^k,$$

we have

$$|h^{k+1}(n) - h_{\lambda^k}(n)| = |F_n(h^k) - F_n(h_{\lambda^k})| \le \alpha\|h^k - h_{\lambda^k}\|. \tag{25}$$

Using this relation and (19), we obtain

$$|h^{k+1}(n)| \le |h^{k+1}(n) - h_{\lambda^k}(n)| + |h_{\lambda^k}(n)| \le \alpha\|h^k - h_{\lambda^k}\| + \overline{N}|\lambda^k - \lambda^*|. \tag{26}$$

We will now derive functions $c_1(\cdot)$ and $c_2(\cdot)$ for which the first and the second relations in (22), respectively, hold. We will then use $c(\gamma) = \max\big[c_1(\gamma), c_2(\gamma)\big]$ in (22). Regarding the first relation in (23), we note that

$$\begin{aligned}\|h^{k+1} - h_{\lambda^{k+1}}\| &= \|F(h^k) - \lambda^k e - F\big(h_{\lambda^{k+1}}\big) + \lambda^{k+1}e\| \\ &\le \|F(h^k) - F\big(h_{\lambda^{k+1}}\big)\| + |\lambda^{k+1} - \lambda^k|\,\|e\| \\ &\le \alpha\|h^k - h_{\lambda^{k+1}}\| + |\lambda^{k+1} - \lambda^k|\,\|e\| \\ &\le \alpha\|h^k - h_{\lambda^k}\| + \alpha\|h_{\lambda^k} - h_{\lambda^{k+1}}\| + |\lambda^{k+1} - \lambda^k|\,\|e\|.\end{aligned} \tag{27}$$

Using the above inequality and (22), (24), and (26), we obtain

$$\begin{aligned}\|h^{k+1} - h_{\lambda^{k+1}}\| &\le \alpha B + \left(\frac{\alpha}{1-\alpha} + 1\right)|\lambda^{k+1} - \lambda^k|\,\|e\| \\ &= \alpha B + \frac{\|e\|\gamma^k}{1-\alpha}|h^{k+1}(n)| \\ &\le \alpha B + \frac{\|e\|\gamma^k}{1-\alpha}\big(\alpha\|h^k - h_{\lambda^k}\| + \overline{N}|\lambda^k - \lambda^*|\big) \\ &\le \alpha B + \frac{\|e\|\gamma^k}{1-\alpha}\left(\alpha B + \frac{\overline{N}B}{\underline{N}}\right) \\ &= c_1(\gamma^k)B,\end{aligned} \tag{28}$$

where $c_1(\cdot)$ is the function

$$c_1(\gamma) = \alpha + \frac{\gamma\|e\|(\alpha + \overline{N}/\underline{N})}{1-\alpha}. \tag{29}$$

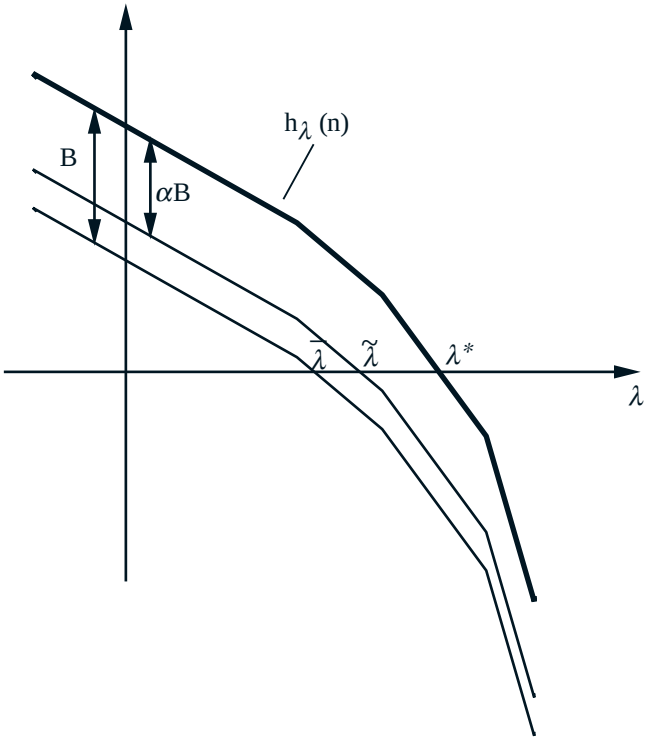

FIG. 3. *Definition of* $\overline{\lambda}$ *and* $\tilde{\lambda}$ *in the proof of Proposition* 1.

Note that if

$$\gamma < \frac{(1-\alpha)^2}{\|e\|(\alpha + \overline{N}/\underline{N})},$$

we have $c_1(\gamma) < 1$.

We now turn to the second relation in (23); that is, we show that

$$|\lambda^{k+1} - \lambda^*| \leq \frac{c_2(\gamma^k)B}{\underline{N}}$$

for an appropriate continuous function $c_2(\gamma)$. Let $\overline{\lambda}$ and $\tilde{\lambda}$ be the unique scalars such that

$$h_{\overline{\lambda}}(n) = B, \qquad h_{\tilde{\lambda}}(n) = \alpha B \tag{30}$$

(see Fig. 3). Also let $\hat{\lambda}$ be the midpoint between $\overline{\lambda}$ and $\tilde{\lambda}$:

$$\hat{\lambda} = \frac{\overline{\lambda} + \tilde{\lambda}}{2}. \tag{31}$$

Note that from (19), we have

$$\frac{(1-\alpha)B}{\overline{N}} \leq \tilde{\lambda} - \overline{\lambda} \leq \frac{(1-\alpha)B}{\underline{N}} \tag{32}$$

and that

$$\frac{\alpha B}{\overline{N}} \leq \lambda^* - \tilde{\lambda} \leq \frac{\alpha B}{\underline{N}},$$

$$\frac{B}{\overline{N}} \leq \lambda^* - \overline{\lambda} \leq \frac{B}{\underline{N}}.$$

From the last three relations, we also obtain

$$\frac{(1+\alpha)B}{2\overline{N}} \leq \lambda^* - \hat{\lambda} \leq \frac{(1+\alpha)B}{2\underline{N}}, \tag{33}$$

$$\frac{(1-\alpha)B}{2\overline{N}} \leq \tilde{\lambda} - \hat{\lambda} \leq \frac{(1-\alpha)B}{2\underline{N}}. \tag{34}$$

We assume that $\lambda^k \leq \lambda^*$; the complementary case where $\lambda^k \geq \lambda^*$ is handled similarly. We distinguish between two cases:

(a) $\lambda^k \leq \hat{\lambda}$,

(b) $\hat{\lambda} < \lambda^k \leq \lambda^*$.

In the case where $\lambda^k \leq \hat{\lambda}$, we have, using (19) and (30)–(32),

$$h_{\lambda^k}(n) \geq h_{\hat{\lambda}}(n) \geq h_{\tilde{\lambda}}(n) + \underline{N}(\tilde{\lambda} - \hat{\lambda}) = \alpha B + \underline{N}(\tilde{\lambda} - \hat{\lambda}) \geq \alpha B + \frac{(1-\alpha)B\underline{N}}{2\overline{N}}. \tag{35}$$

On the other hand, from (22) and (25), we have $|h^{k+1}(n) - h_{\lambda^k}(n)| \leq \alpha B$ so that

$$h^{k+1}(n) \geq h_{\lambda^k}(n) - \alpha B. \tag{36}$$

By combining (35) and (36), we obtain

$$h^{k+1}(n) \geq \frac{(1-\alpha)B}{2\overline{N}^2}.$$

We now have, using the above equation,

$$\lambda^* - \lambda^{k+1} = \lambda^* - \lambda^k - \gamma^k h^{k+1}(n) \leq \frac{B}{\underline{N}} - \frac{\gamma^k(1-\alpha)B\underline{N}}{2\overline{N}} = \frac{B}{\underline{N}^2}\left(1 - \frac{\gamma^k(1-\alpha)\underline{N}}{2\overline{N}}\right), \tag{37}$$

and we also have, using (25), (22), and (19)

$$\lambda^* - \lambda^{k+1} = \lambda^* - \lambda^k - \gamma^k h^{k+1}(n) \geq \lambda^* - \lambda^k - \gamma^k\big(h_{\lambda_k}(n) + \alpha B\big) \geq (1 - \gamma^k\overline{N})(\lambda^* - \lambda^k) - \gamma^k\alpha B. \tag{38}$$

It can be seen now from (38) that for $\gamma^k \in (0, 1/\overline{N}]$, we have $\lambda^* - \lambda^{k+1} \geq -\gamma^k\alpha B$, and it follows using also (37) that

$$|\lambda^* - \lambda^{k+1}| \leq \frac{c_2(\gamma^k)B}{\underline{N}},$$

where $c_2(\cdot)$ is the continuous function

$$c_2(\gamma) = \max\left[1 - \frac{\gamma(1-\alpha)\underline{N}^2}{2\overline{N}},\ \gamma\alpha\underline{N}\right].$$

Since there exists a threshold value $\overline{\gamma} > 0$ such that the continuous function $c_2(\gamma)$ satisfies $0 < c(\gamma) < 1$ for all $\gamma \in (0, \overline{\gamma}]$, the desired relation (23) is proved in the case $\lambda^k \leq \hat{\lambda}$.

In the case where $\hat{\lambda} < \lambda^k \leq \lambda^*$, there are two possibilities.

(1) $h^{k+1}(n) \geq 0$. Then $\lambda^k \leq \lambda^{k+1}$, and by also using (33), we have

$$\lambda^* \leq \hat{\lambda} + \frac{(1+\alpha)B}{2\underline{N}} \leq \lambda^k + \frac{(1+\alpha)B}{2\underline{N}} \leq \lambda^{k+1} + \frac{(1+\alpha)B}{2\underline{N}}. \tag{39}$$

Furthermore, from (22) and (26), we have

$$\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n) \leq \lambda^* + \gamma^k \left(\alpha B + \frac{\overline{N}B}{\underline{N}} \right).$$

Thus, by choosing γ^k sufficiently small, we can guarantee that

$$\lambda^{k+1} \leq \lambda^* + \frac{(1+\alpha)B}{2\underline{N}}. \tag{40}$$

From (39) and (40), it follows that for γ^k less than some positive constant, we have

$$|\lambda^{k+1} - \lambda^*| \leq \frac{(1+\alpha)B}{2\underline{N}},$$

proving the second relation in (23), with $c_2(\gamma) = (1+\alpha)/2$.

(2) $h^{k+1}(n) < 0$. In this case, since from (22) and (25) we have

$$h_{\lambda^k}(n) \leq h^{k+1}(n) + \alpha B \leq \alpha B, \tag{41}$$

and since $h_{\tilde{\lambda}}(n) = \alpha B$ and $h_\lambda(n)$ is monotonically decreasing in λ, it follows that $\tilde{\lambda} \leq \lambda^k$. Since $\lambda^k \leq \lambda^*$, we also have $0 \leq h_{\lambda^k}(n) \leq \alpha B$, so that by using (41) and the fact $h_{\lambda^k}(n) \geq 0$, we obtain $|h^{k+1}(n)| \leq \alpha B$ and

$$|\gamma^k h^{k+1}(n)| \leq \gamma^k \alpha B.$$

By choosing

$$\gamma^k \in \left(0, \frac{1-\alpha}{2\alpha\overline{N}} \right], \tag{42}$$

the above inequality, together with (34), yields

$$|\gamma^k h^{k+1}(n)| \leq \frac{(1-\alpha)B}{2\overline{N}} \leq \tilde{\lambda} - \hat{\lambda} \leq \lambda^k - \hat{\lambda}.$$

Thus, we have

$$\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n) \geq \hat{\lambda},$$

and from (33), using also the fact $\lambda^{k+1} \leq \lambda^k \leq \lambda^*$, we obtain for γ^k satisfying (42),

$$|\lambda^{k+1} - \lambda^*| \leq \frac{(1+\alpha)B}{2\underline{N}},$$

proving the second relation in (23) for the case $h^{k+1}(n) < 0$ as well.

Thus, (23) holds with $c(\cdot)$ given by

$$c(\gamma) = \max \left[\alpha + \frac{\gamma \|e\| (\alpha + \overline{N}/\underline{N})}{1-\alpha},\ 1 - \frac{\gamma(1-\alpha)\underline{N}^2}{2\overline{N}},\ \gamma\alpha\underline{N},\ \frac{1+\alpha}{2} \right]. \qquad \square$$

3. Convergence analysis of the Gauss–Seidel version. In this section, we prove the result of Proposition 1 for the Gauss–Seidel version of the method, given by

$$h^{k+1}(i) = G_i(h^k, \lambda^k), \qquad i = 1, \ldots, n, \tag{43}$$

$$\lambda^{k+1} = \lambda^k + \gamma^k h^{k+1}(n), \tag{44}$$

where the components of the mapping $G = (G_1, \ldots, G_n)$ are given by

$$G_1(h, \lambda) = \min_{u \in U(1)} \left[g(1, u) + \sum_{j=1}^{n-1} p_{1j}(u) h(j) \right] - \lambda, \tag{45}$$

(46)

$$G_i(h, \lambda) = \min_{u \in U(i)} \left[g(i, u) + \sum_{j=1}^{i-1} p_{ij}(u) G_j(h, \lambda) + \sum_{j=i}^{n-1} p_{ij}(u) h(j) \right] - \lambda, \qquad i = 2, \ldots, n.$$

The proof of Proposition 1 essentially carries through with the aid of the following result.

PROPOSITION 2. *The mapping G of* (45) *and* (46) *satisfies for all $h \in \Re^n$, $\overline{h} \in \Re^n$, $\lambda \in \Re$, and $\overline{\lambda} \in \Re$:*

$$\frac{|G_i(h, \lambda) - G_i(\overline{h}, \overline{\lambda})|}{v_i} \leq \alpha \|h - \overline{h}\| + \delta_i |\lambda - \overline{\lambda}| \qquad \forall\ i = 1, \ldots, n, \tag{47}$$

where α is the contraction modulus of F, $v_1, \ldots, v_n$ are the weights of the sup-norm $\|\cdot\|$ with respect to which F is a contraction (cf. (17)*), and $\delta_1, \ldots, \delta_n$ are defined recursively by*

$$\delta_1 = \frac{1}{v_1}, \qquad \delta_i = \frac{1 + \max_{j=1,\ldots,i-1} \delta_j}{v_i}, \quad i = 2, \ldots, n. \tag{48}$$

In particular, by taking the maximum over i in (47)*, we obtain*

$$\|G(h, \lambda) - G(\overline{h}, \overline{\lambda})\| \leq \alpha \|h - \overline{h}\| + \delta |\lambda - \overline{\lambda}| \qquad \forall\ h \in \Re^n,\ \overline{h} \in \Re^n,\ \lambda \in \Re,\ \overline{\lambda} \in \Re, \tag{49}$$

where

$$\delta = \max_{i=1,\ldots,n} \delta_i.$$

Proof. We prove (47) by induction. For the case where $i = 1$, we have from the contraction property of the mapping F (cf. (17)):

$$\frac{|G_1(h, \lambda) - G_1(\overline{h}, \lambda)|}{v_1} \leq \alpha \max_{i=1,\ldots,n} \frac{|h(i) - \overline{h}(i)|}{v_i} = \alpha \|h - \overline{h}\|.$$

Therefore,

$$\begin{aligned} \frac{G_1(h, \lambda)}{v_1} &\leq \frac{G_1(\overline{h}, \lambda)}{v_1} + \alpha \|h - \overline{h}\| \\ &\leq \frac{G_1(\overline{h}, \overline{\lambda})}{v_1} + \alpha \|h - \overline{h}\| + \frac{|\lambda - \overline{\lambda}|}{v_1}. \end{aligned}$$

Similarly, we obtain

$$\frac{G_1(\overline{h},\overline{\lambda})}{v_1} \le \frac{G_1(h,\lambda)}{v_1} + \alpha\|h-\overline{h}\| + \frac{|\lambda-\overline{\lambda}|}{v_1}.$$

By combining the last two relations, we see that

$$\frac{|G_1(h,\lambda)-G_1(\overline{h},\overline{\lambda})|}{v_1} \le \alpha\|h-\overline{h}\| + \delta_1|\lambda-\overline{\lambda}|,$$

so that (47) is proved for $i=1$.

Assume that (47) holds for $i=1,\dots,m-1$. We will show that it holds for $i=m$. We have from the contraction property of the mapping F and the induction hypothesis

$$\begin{aligned}
&\frac{|G_m(h,\lambda)-G_m(\overline{h},\lambda)|}{v_m} \\
&\quad\le \alpha\max\left\{\max_{i=1,\dots,m-1}\frac{|G_i(h,\lambda)-G_i(\overline{h},\lambda)|}{v_i}, \max_{i=m,\dots,n}\frac{|h(i)-\overline{h}(i)|}{v_i}\right\} \\
&\quad\le \alpha\|h-\overline{h}\|.
\end{aligned}$$

Using this relation and the induction hypothesis, we obtain

$$\begin{aligned}
\frac{G_m(h,\lambda)}{v_m} &\le \frac{G_m(\overline{h},\lambda)}{v_m} + \alpha\|h-\overline{h}\| \\
&= \frac{1}{v_m}\min_{u\in U(m)}\left[g(m,u)+\sum_{j=1}^{m-1}p_{mj}(u)G_j(\overline{h},\lambda)+\sum_{j=m}^{n-1}p_{mj}(u)\overline{h}(j)\right] \\
&\quad -\frac{\lambda}{v_m}+\alpha\|h-\overline{h}\| \\
&\le \frac{1}{v_m}\min_{u\in U(m)}\left[g(m,u)+\sum_{j=1}^{m-1}p_{mj}(u)G_j(\overline{h},\overline{\lambda})+\sum_{j=m}^{n-1}p_{mj}(u)\overline{h}(j)\right]-\frac{\overline{\lambda}}{v_m} \\
&\quad +\frac{|\lambda-\overline{\lambda}|}{v_m}+\max_{j=1,\dots,m-1}\delta_j\frac{|\lambda-\overline{\lambda}|}{v_m}+\alpha\|h-\overline{h}\| \\
&= \frac{G_m(\overline{h},\overline{\lambda})}{v_m}+\delta_m|\lambda-\overline{\lambda}|+\alpha\|h-\overline{h}\|.
\end{aligned}$$

Similarly, we obtain

$$\frac{G_m(\overline{h},\overline{\lambda})}{v_m} \le \frac{G_m(h,\lambda)}{v_m} + \delta_m|\lambda-\overline{\lambda}| + \alpha\|h-\overline{h}\|,$$

thus proving (47) for $i=m$. This completes the induction. □

Note that Proposition 1 implies that for any λ, $G(\cdot,\lambda)$ is a weighted sup-norm contraction when viewed as a function of h. It can be easily verified that

$$h_\lambda = G(h_\lambda,\lambda) \qquad \forall\ \lambda\in\Re,$$

so it follows that for all λ, the mapping $G(\cdot,\lambda)$ has h_λ as its unique fixed point. The following result proves convergence of the Gauss–Seidel method and parallels Proposition 1.

TABLE 1

n	Sparsity	STANDARD	SSP-JACOBI	SSP-Gauss–Seidel
10	0.5	16	39	40
20	0.5	9	39	75
30	0.5	9	48	105
40	0.5	8	46	55
50	0.5	8	56	90
10	0.1	674	727	185
20	0.1	202	203	160
30	0.1	38	66	130
40	0.1	36	77	75
50	0.1	21	63	110
10	0.05	114	294	70
20	0.05	131	145	100
30	0.05	49	53	235
40	0.05	259	226	205
50	0.05	313	313	325

TABLE 2

n	STANDARD	SSP-JACOBI	SSP-Gauss–Seidel
10	211	211	180
20	2658	2658	2070
30	29638	29647	20615
40	286550	286765	222855
50	13219	13217	9035

PROPOSITION 3. *There exists a positive scalar $\overline{\gamma}$ such that if*

$$\underline{\gamma} \leq \gamma^k \leq \overline{\gamma}$$

for some positive scalar $\underline{\gamma}$ and all k, the sequence (h^k, λ^k) generated by the Gauss–Seidel iteration (43), (44) converges to $(h_{\lambda^*}, \lambda^*)$ at the rate of a geometric progression.

Proof. The proof is essentially identical to the one of Proposition 1. The only difference is that the three relations (27), (28), and (29) must be modified to involve the mapping G and to make use of Proposition 2. In particular, (27) becomes

$$\|h^{k+1} - h_{\lambda^{k+1}}\| \leq \alpha\|h^k - h_{\lambda^k}\| + \alpha\|h_{\lambda^k} - h_{\lambda^{k+1}}\| + \delta|\lambda^{k+1} - \lambda^k|,$$

and (28) becomes

$$\|h^{k+1} - h_{\lambda^{k+1}}\| \leq c_1(\gamma^k)B,$$

where the function $c_1(\cdot)$ of (29) is now given by

$$c_1(\gamma) = \alpha + \gamma\left(\frac{\alpha\|e\|}{1-\alpha} + \delta\right)\left(\alpha + \frac{\overline{N}}{\underline{N}}\right).$$

The remainder of the proof goes through with no modification. □

4. Implementation and experimentation. In this section we describe some of our computational experience with the standard method (9)–(10) and with the new Jacobi and Gauss–Seidel methods. The latter methods were implemented with an adaptive stepsize rule of the form $\gamma^k = m(\hat{k})\gamma$ (cf. (15)), using an initial stepsize γ equal to 1. We used the function $m(\hat{k})$ of (16a) for the test results of Tables 1–3 and the function $m(\hat{k})$ of (16b) for the test results of Table 4. The projection

TABLE 3

n	STANDARD	SSP-JACOBI	SSP-Gauss–Seidel
10	121	119	80
20	826	825	545
30	18020	18026	13465
40	2186	2186	1360
50	5942	5941	4770
75	7978	7984	5000
100	9035	9028	6880
125	10306	10323	7440
150	9011	9015	6870

TABLE 4

n	STANDARD	SSP-JACOBI	SSP-Gauss–Seidel
250	939	940	420
500	4724	4725	470
750	1257	1257	740
1000	710	711	1040
1250	1693	1693	1425
1500	2870	2870	1890
1750	5605	5609	4230
2000	4691	4693	3180

scheme of (11)–(14) was also used. To obtain error bounds on which to project in the Gauss–Seidel method, we performed one Jacobi iteration following nine consecutive Gauss–Seidel iterations. Each Jacobi iteration yielded an upper and a lower bound for λ^*, and the λ-iterate obtained by each iteration was projected on the interval of the best upper and lower bounds obtained so far. For each problem, the three methods were initialized with $h = 0$ and (for the case of the new methods) $\lambda = n/2$. Note that because of the device of projection on the error bound range, the initial choice of λ is not critical.

Our computational results with randomly generated problems are summarized in Tables 1–4 for the three methods labeled STANDARD (which is the known iteration (9)–(10)), SSP-JACOBI (which is the Jacobi version of the new method (7)–(8)), and SSP-Gauss–Seidel (which is the Gauss–Seidel version of the new method (43)–(44)). Let us describe how the test problems were generated. Regarding cost structure, in all problems and for each pair (i, u), the one-stage cost at state i was randomly selected from the range $(0, n)$ according to a uniform distribution. Regarding the transition probabilities, in all the problems, we specified the structure of the transition probability graph by specifying for each state-control pair (i, u), according to some (possibly random) rule, the states j for which the transition probability $p_{ij}(u)$ is nonzero. We then generated each of the nonzero transition probabilities by randomly selecting a corresponding number from the interval $(0, 1)$ according to uniform probability distribution, and by normalizing so that $\sum_{j=1}^{n} p_{ij}(u) = 1$ for all pairs (i, u). The test problems were generated as follows.

(1) Problems of Table 1. Here there is only one control available at each state. The sparsity of the transition probability graph is controlled by a parameter $q \in (0, 1)$. In particular, each possible transition probability is selected to be nonzero with a given probability q. We used sparsity parameters $q = 0.5$, $q = 0.1$, and $q = 0.05$ in our tests.

(2) Problems of Table 2. Here also there is only one control available at each state. At states i with $1 < i < n$, the nonzero transition probabilities are the ones to

the states $i-1$, i, and $i+1$. At state 1 the nonzero transition probabilities are to states 1 and 2, and at state n the nonzero transition probabilities are to states $n-1$ and n. This type of transition probability graph arises in queueing systems.

(3) Problems of Table 3. Here there are two controls available at each state, call them u_1 and u_2. Under u_1, the transition probabilities are specified in the same way as for the problems of Table 2. Under u_2, at each state i with $1 < i < n$, the nonzero transition probabilities are the ones to the states $i-1$ and $i+1$. At state 1 the only nonzero transition probabilities are the ones to the states 1 and 2, and at state n the only nonzero transition probabilities are the ones to the states $n-1$ and n.

(4) Problems of Table 4. Here there are three controls available at each state, call them u_1, u_2, and u_3. Under u_1, the transition probabilities are specified in the same way as for the problems of Table 2. Under u_2, at each state i with $1 < i < n-10$, the nonzero transition probabilities are the ones to the states $i-1$ and $i+10$. At state 1 the only nonzero transition probabilities are the ones to the states 1 and 11, and at states $i = n-10, n-9, \ldots, n$ the only nonzero transition probabilities are the ones to the states $i-1$ and n. Under u_3, at each state i with $10 < i < n$, the nonzero transition probabilities are the ones to the states $i-10$ and $i+1$. At states $i = 1, \ldots, 10$, the only nonzero transition probabilities are the ones to the states 1 and $i+1$, and at state n the only nonzero transition probabilities are the ones to the states $n-10$ and n.

Tables 1–4 give the number of iterations required by each method for the difference between the upper and lower bounds to become smaller than 10^{-3}. Each entry of the tables represents the average of two problems. We should note here that the number of iterations varies a great deal from one problem to another, so the variance of the number of iterations for a given type of problem is very large. For example, one of the two problems in the fourth entry of Table 2 is extremely difficult and requires a much larger number of iterations than the other. However, it is generally true that if a problem is difficult for one method (requires a lot of iterations), it is also difficult for all the other methods.

It can be seen that the problems of Table 1 are generally much easier than the problems of Tables 2–4. Generally, it appears that these problems become more difficult as the sparsity of the transition probability graph increases. On some of these problems (generally the easier ones), the standard method performs extremely well and much better than the new methods. This is probably due to the need for stepsize selection in the new methods. The adaptive stepsize rule that we used generally works well, but on occasion may end up with a stepsize that is either too large or too small for optimal performance. We believe that the subject of appropriate stepsize selection method is a potential topic for theoretical or empirical research.

On the more difficult problems of Tables 2 and 3, the Gauss–Seidel version of the new method is uniformly faster than the other methods. In fact, the Gauss–Seidel method has substantially outperformed the other methods on every single problem with the queueing structure that we tried. The Jacobi version of the new method performs comparably to the standard method on the problems of Tables 2 and 3. What happens here is that for the problems of Tables 2 and 3, the difference between the iterations (7)–(8) and (9)–(10) are minor, particularly when the number of states is large (see the discussion following (9)–(10)).

For the larger problems of Table 4, again the Jacobi version of the new method performs comparably to the standard method. The Gauss–Seidel version of the new method is generally faster than the other methods, but the factor of superiority is problem dependent and its variance is substantial.

5. Conclusions. The methods of this paper were derived by exploiting the connection between average cost and stochastic shortest path problems. We developed a new value iteration method that involves the same type of weighted sup-norm contraction that arises in stochastic shortest path problems. This method is the first, to our knowledge, that admits a convergent Gauss–Seidel implementation. We also believe that the weighted sup-norm contraction property inherent in our method is likely to prove useful in other related contexts.

REFERENCES

[ABB97] J. Abounadi, D. P. Bertsekas, and V. Borkar, *Q-Learning Algorithms for the Average Cost Markovian Decision Problem*, in preparation, 1997.

[BeT89] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[Ber95] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vols.* I *and* II, Athena Scientific, Belmont, MA, 1995.

[Odo69] A. R. Odoni *On finding the maximal gain for Markov decision processes*, Oper. Res., 17 (1969), pp. 857–860.

[PBW79] J. L. Popyack, R. L. Brown, and C. C. White, III, *Discrete versions of an algorithm due to Varaiya*, IEEE Trans. Automat. Control, 24 (1969), pp. 503–504.

[Pla77] L. Platzman, *Improved conditions for convergence in undiscounted Markov renewal programming*, Oper. Res., 25 (1977), pp. 529–533.

[Put94] M. L. Puterman, *Markovian Decision Problems*, Wiley, New York, 1994.

[Sch71] P. J. Schweitzer, *Iterative solution of the functional equations of undiscounted Markov renewal programming*, J. Math. Anal. Appl., 34 (1971), pp. 495–501.

[Tse90] P. Tseng, *Solving H-horizon, stationary Markov decision problems in time proportional to* $\log(H)$, Oper. Res. Lett., 9, (1990), pp. 287–297.

[Var78] P. P. Varaiya, *Optimal and suboptimal stationary controls of Markov chains*, IEEE Trans. Automat. Control, AC-23 (1978), pp. 388–394.

[Whi63] D. J. White, *Dynamic programming, Markov chains, and the method of successive approximations*, J. Math. Anal. Appl., 6 (1963), pp. 373–376.